International Journal of Applied Earth Observation and Geoinformation 146 (2026) 105161

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

# Enhancing the interpretability of port economic modeling via implicit spatial relationship discovery

Xiaoli Cai [a], Sheng Wu [a,*], Peixiao Wang [b,c,*], Hengcai Zhang [b,c], Shifen Cheng [b,c], Feng Lu [b,c]

[a] Academy of Digital China (Fujian), Fuzhou University, Fuzhou 350002, China
[b] State Key Laboratory of Geographic Information Science and Technology, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
[c] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

*Keywords:*
Port economy
PortCity2Vec
Spatial XGBoost
GeoShapley
AIS data

ABSTRACT

Modeling the driving mechanisms of economic activities among port cities helps reveal their interactions and spatial spillover effects, which is crucial for promoting coordinated regional economic development. Current mainstream models of these driving mechanisms are mostly based on machine learning integrated with the SHAP method, but often neglect spatial dependencies between samples—especially the implicit spatial relationships underlying port city economic activities. In recent years, AIS data has become an important tool for uncovering these implicit spatial relationships. Therefore, we propose the PortCity2Vec framework, based on AIS data and embedding representation learning, to explicitly capture implicit spatial relationships among port cities. Furthermore, we develop a spatial XGBoost model integrated with GeoShapley to incorporate these implicit spatial relationships, thereby revealing the driving mechanisms behind socioeconomic indicators and quantifying the core roles of spatial relationships and geographic features in the port economy. The results show that: (1) Implicit economic interactions among port cities extend beyond physical adjacency, indicating that port economy is influenced not only by physical proximity but also by connections within an implicit spatial structure; (2) Introducing socioeconomic indicators and geographic features of nearby port cities within the implicit spatial structure improves model accuracy, increasing $R^2$ from 0.6554 to 0.8541; (3) The port economy correlates positively with cargo turnover and grain crop output, but negatively with forestry and meat output. These findings highlight the key roles of economic and transport intensity and reveal resource allocation gaps. (4) Embeddings of implicit spatial relationships and geographic features effectively capture regional potential economic connections and marginal contributions that traditional models struggle to identify, thereby enhancing both the performance and interpretability of the model.

## 1. Introduction

With the advancement of the global sustainable development agenda, ports—key nodes for resource aggregation, factor mobility, and regional collaboration—have attracted widespread attention for both their economic functions and sustainable development potential (Zhou et al., 2025). We define a port city as an administrative urban area whose economy and infrastructure are significantly shaped by port-centered activities, serving as both a hub for maritime trade and a platform for industrial and logistical development. This includes both coastal and inland cities hosting major ports along sea or river shipping routes. Systematic research on the port economy contributes to optimizing port resource allocation and enhancing the competitiveness of port cities (Peng et al., 2018), while also supporting green growth and

*Abbreviations:* QZF, Quanzhou; FZ, Fuzhou; XM, Xiamen; PT, Putian; ZZ, Zhangzhou; ND, Ningde; WZG, Wuzhou; GG, Guigang; NN, Nanning; CM, Chengmai; FCG, Fangchenggang; QZG, Qinzhou; BH, Beihai; DZ, Danzhou; HK, Haikou; SY, Sanya; DF, Dongfang; GZ, Guangzhou; ST, Shantou; ZH, Zhuhai; SZG, Shenzhen; ZJG, Zhanjiang; MM, Maiming; YJ, Yangjiang; HZG, Huizhou; SW, Shanwei; DG, Dongguan; ZQ, Zhaoqing; FS, Foshan; ZSG, Zhongshan; JM, Jiangmen; YF, Yunfu; CZG, Chaozhou; JY, Jieyang; HZU, Huzhou; HZ, Hangzhou; ZSZ, Zhoushan; NB, Ningbo; WZZ, Wenzhou; JX, Jiaxing; SH, Shanghai; LYG, Liangyungang; NJ, Nanjing; ZJJ, Zhenjiang; SZJ, Suzhou; NT, Nantong; WX, Wuxi; YZ, Yangzhou; TZ, Taizhou; CZJ, Changzhou; HA, Huaian; YC, Yancheng; XZ, Xuzhou.

* Corresponding authors..
*E-mail addresses:* wusheng@fzu.edu.cn (S. Wu), wpx@lreis.ac.cn (P. Wang).

https://doi.org/10.1016/j.jag.2026.105161
Received 3 June 2025; Received in revised form 23 December 2025; Accepted 6 February 2026
1569-8432/

coordinated development of regional economies (Zhou and Yuen, 2024). According to the "Port Economic Development Report of Chinese Port Cities" released by the Ministry of Transport, "Port Economy" refers to the total economic activities centered around ports, with port cities serving as key carriers (Zhang et al., 2024c). As port activities become increasingly frequent, the role of the port economy as a hub in regional economic integration and intercity shipping connectivity has become more prominent. In this context, gaining a deep understanding of the mechanisms driving the port economy development is critical to advancing regional coordination and integration (Zhang et al., 2024b).

In the era of big data, research on the driving mechanisms of the port economy is gradually shifting from the traditional paradigms based on statistical and econometric methods (Henríquez et al., 2022; Yen et al., 2023) toward more intelligent and data-driven analytical approaches (Peng et al., 2020). Early research primarily relied on econometric models with linear assumptions (Li et al., 2023). For example, Wang and Tan (2019) employed cointegration analysis, an error correction model, and the Granger causality test to construct a dynamic model for their study, revealing that port cargo throughput significantly influences regional economic performance. Zhao et al. (2020) applied the entropy-weighted TOPSIS method and the spatial Durbin model to quantitatively assess port intensity and its spatial spillover effects on urban economic growth. Lin et al. (2024) developed a dynamic spatial Durbin model to examine the coordinated development mechanisms among port cities in Guangdong from a digital economy perspective. Although these studies have made some progress in revealing the driving mechanisms of the port economy, most of them rely on linear models. This makes it difficult to capture the nonlinear relationships and complex dynamic interactions within the system. As the port economy evolves towards networking and intelligence, its driving mechanisms increasingly involve multi-factor coupling and multi-scale interactions. Traditional approaches are thus limited in uncovering the internal logic and evolutionary pathways of port economic systems, highlighting the need for more flexible and robust research methodologies.

In recent years, the development of machine learning has brought new breakthroughs to the study of port economy driving mechanisms (Liu, 2024). Machine learning techniques have been widely applied in port economy research, demonstrating superior performance compared to traditional statistical methods in areas such as port economic modeling (Chen et al., 2024), port traffic forecasting (Filom et al., 2022; Lim et al., 2021), and transportation efficiency optimization (Durán et al., 2024). However, current machine learning models are often regarded as "black boxes" due to their limited interpretability, which constrains their applicability in analyzing economic driving mechanisms (Park and Yang, 2022). To overcome this limitation, researchers have introduced SHapley Additive exPlanations (SHAP) methods to improve model transparency by uncovering the internal mechanisms of these models (Dwivedi et al., 2023). Although SHAP has been widely applied in fields such as finance (Černevičienė and Kabašinskas, 2024; Ueda et al., 2024; Zhao et al., 2020), healthcare (Bhattarai et al., 2024; Ghebrehiwet et al., 2024; Nohara et al., 2022), and image classification (Ishikawa et al., 2023), conventional SHAP methods remain limited in their ability to capture spatial correlations in geographic analysis. This limitation stems from the assumption underlying most machine learning models and SHAP-based interpretations—that data are independent and identically distributed (i.i.d.) (Hu et al., 2024). In geographic contexts, however, this assumption is often violated due to significant spatial dependence among samples. In the case of the port economy, spatial dependence frequently arises from indirect and implicit economic interactions rather than direct physical adjacency, which further complicates the analysis of driving mechanisms.

Building on the limitations of traditional spatial analyses in capturing complex economic interactions, we define implicit spatial relationships as latent or indirect connections among port cities that are not determined by physical adjacency or administrative boundaries. Unlike explicit spatial relationships, which depend on direct adjacency or formal institutional ties, implicit spatial relationships arise from actual interactions such as the flow of goods, movement of people, or industrial complementarities. These relationships often reflect complex, higher-order network patterns that traditional spatial metrics cannot fully capture.

To model these hidden relationships, we leverage Automatic Identification System (AIS) data, which records the berthing of ships at ports and their movements (Peng et al., 2020; Yang et al., 2019). In addition to monitoring maritime traffic (Liu et al., 2020), AIS data offers a valuable means of capturing the functional and economic linkages among port cities (Cong et al., 2020), providing a data-rich foundation for uncovering implicit spatial dependencies in global shipping networks (Feng et al., 2024; Liu et al., 2024; Peng et al., 2023). We propose a novel PortCity2Vec framework that uses AIS data to learn high-dimensional embeddings of port cities based on co-occurrence probabilities within the shipping network. These embeddings are then used as input features in a spatial XGBoost model integrated with interpretability techniques to uncover implicit spatial and economic interaction patterns. This combined approach captures implicit spatial relationships among port cities, uncovering functional linkages between non-physically adjacent cities that are often overlooked in conventional, explicit models. As a result, modeling implicit spatial relationships provides a more nuanced and powerful perspective on economic interactions and spatial spillovers across port cities. Specifically, the contributions of this study are as follows:

(1) We propose a novel PortCity2Vec framework that uses AIS data and embedding representation learning to reveal implicit spatial relationships among port cities and uncover their underlying spatial dynamics through the shipping mobility network.

(2) We construct a spatial XGBoost model integrated with GeoShapley that incorporates implicit spatial relationships, reveals the impact mechanism of socioeconomic indicators on the port economy, and systematically quantifies the core driving roles of implicit spatial structures and geographic factors.

(3) Using port cities in southeastern coastal China as a case study, we uncover clustering patterns beyond physical adjacency. Furthermore, by incorporating neighboring port city features derived from the implicit spatial structure and geographic features, we highlight the role of implicit spatial features in economic connections and clarify the spatial spillover effects and cross-regional interactions in the port economy development.

## 2. Study area and data sources

### 2.1. Study area

The southeastern coastal region of China stands as one of the nation's most economically dynamic areas, encompassing major economic zones such as the Yangtze River Delta and the Pearl River Delta. Characterized by a high density of ports and a sophisticated shipping mobility network, this region plays a central role in supporting regional industrial chains, trade circulation, and logistics systems. Analyzing economic linkages among port cities helps reveal their complex interactions and supports the coordinated development and integration of the regional economy.

To align AIS-recorded port activity with administrative boundaries, we define port cities as administrative units that satisfy the following criteria during the study year: (1) at least one port within the city's administrative boundary exhibits regular vessel berthing activities recorded in the AIS data; (2) the city reports non-zero port cargo throughput in the corresponding official statistical yearbooks. This definition ensures consistency between AIS-observed shipping activities and officially recognized port economic functions.

### 2.2. Data sources and preprocessing

We utilize 2022 ship AIS data provided by Hifleet, an online platform specializing in global shipping and vessel dynamics tracking. Hifleet widely serves the shipping, logistics and trade industries by providing critical information, including real-time and historical vessel positions, route trajectories, and port activity dynamics.

AIS (Automatic Identification System) data are primarily used for ship identification, location tracking and navigation status monitoring, serving as a fundamental resource for studying maritime transport networks (Xu et al., 2024). However, raw AIS data often contain redundancies, missing values and outliers, requiring thorough preprocessing to enhance data quality and ensure the accuracy of subsequent analyses.

We processed AIS data through the following steps to construct a consistent city-level shipping network: (1) Data cleaning: Records with missing key fields—such as MMSI, berthing time, departure time, or port name—were removed. Duplicate records were eliminated. We retained only essential fields to improve data quality. (2) Study area filtering: The dataset was restricted to shipping activities along China's southeast coast. Records outside the study area were excluded. (3) City-level belonging: AIS records sometimes lack explicit city identifiers. Records with missing or ambiguous city belonging were mapped to officially published port administration documents specifying port-city jurisdiction. The resulting city-level AIS berthing aggregates were compared with official port statistics as a consistency check, ensuring alignment between AIS observations and recognized port-city relationships. (4) Construction of ship berthing trajectory sequences: The AIS data were sorted by MMSI and berthing time to construct berthing trajectory sequences, which reflect shipping patterns among port cities. Table 1 summarizes the final set of retained AIS data fields.

The selection of socioeconomic variables was informed by both established literature and the specific analytical goals of this study. Prior research commonly employs indicators such as port cargo throughput, regional GDP, cargo transportation and turnover volumes as proxies for port economic performance and regional development (Cong et al., 2020; Li et al., 2019; Zhang et al., 2024c). To better capture structural economic differences across cities, we also incorporated sector-specific indicators including outputs of agriculture, forestry, animal husbandry, and fisheries. These reflect how dominant industries affect port-related trade flows and logistics demand. Specifically, agricultural productivity—indicators including crop, vegetable and fruit, poultry egg, meat, and aquatic product outputs—serve as direct indicators of regional agricultural trade potential. These variables highlight the role of agriculture in driving trade flows and shaping port cargo throughput. Moreover, total and industrial electricity consumption were included as proxies for production intensity, energy demand, and overall economic activity. These indicators help characterize the scale of industrial operations and their impact on port logistics demand. Overall, these variables offer a comprehensive perspective on the economic and logistical underpinnings of cargo throughput and inter-port connectivity. Although port infrastructure indicators such as berth number and terminal capacity are important determinants of port performance, they were not included in this study for two reasons. First, these indicators are often highly correlated with port cargo throughput, which may introduce endogeneity and obscure the interpretation of model results. Second, consistent and comparable infrastructure statistics at the city level are not publicly available for all port cities. Therefore, this study focuses on socioeconomic and flow-based indicators that reflect demand-side and structural economic characteristics, which are less likely to introduce endogeneity and offer better comparability across port cities. Based on the socioeconomic variables presented above, and as shown in Table 2, the data were sourced primarily from the 2023 Statistical Yearbook of port cities. To ensure data quality and model interpretability, we conducted thorough data cleaning and filtering. Port cargo throughput was selected as the dependent variable to represent the port's transportation capacity and economic vitality, while other variables served as explanatory variables that have an influence on port performance. (See Fig. 1.).

**Table 1**
AIS data field description.

| Field | Description | Example data |
|---|---|---|
| MMSI | Unique Ship ID | 413,800,531 |
| atb | Berthing Time | 2022/6/14 11:31:12 |
| atd | Departure Time | 2022/6/17 13:19:43 |
| portname | Docking Port | Da Chan Bay Terminal One |
| city | Docking Port City | Shenzhen |

## 3. Methods

As shown in Fig. 2, the technical framework of this study comprises three core components: (1) identifying implicit spatial relationships based on PortCity2Vec; (2) constructing a spatial XGBoost model that incorporates these relationships; and (3) applying GeoShapley to analyze model interpretability while accounting for spatial dependencies. PortCity2Vec generates high-dimensional embeddings of port cities based on the shipping mobility network, making the implicit spatial relationships explicit. Next, we build a spatial XGBoost model that incorporates implicit spatial relationships by leveraging the similarity of port city embeddings. The model integrates independent features of each port city and captures implicit spatial dependencies through embeddings to reveal potential economic connections. Finally, we apply GeoShapley to interpret the XGBoost model and systematically quantify the driving mechanisms behind the port economy, accounting for both implicit spatial relationships and geographic features.

### *3.1. Discovery of implicit spatial relationships based on PortCity2Vec*

Identifying spatial relationships is fundamental to understanding the complex interaction mechanisms of the port economy. Traditional approaches largely rely on physical adjacency, which fails to capture implicit spatial relationships among port cities. Inspired by Word2Vec in natural language processing—which maps discrete words into

**Table 2**
Description of port city socioeconomic indicator variables.

| Variables | Indicator | Description |
|---|---|---|
| *Y* | port cargo throughput | the transportation capacity of a region's ports. |
| *rGDP* | regional GDP | total economic value of all production activities within a region |
| *CTRV* | cargo transportation volume | total volume of cargo transported within a region |
| *CTUV* | cargo turnover volume | total volume of cargo flowing through a region |
| *TEC* | total electricity consumption | total electricity usage by total sectors within a region |
| *IEC* | industrial electricity consumption | total electricity usage by industrial sector within a region |
| *TOV* | total output value of agriculture, forestry, animal husbandry, and fishery industry | reflects the overall output value of agriculture, forestry, animal husbandry, and fishery industry within a region |
| *AOV* | agricultural output value | reflects agricultural output value |
| *FOV* | forestry output value | reflects forestry output value output value |
| *AHOV* | animal husbandry output value | reflects animal husbandry output value |
| *FIOV* | fishery output value | reflects agriculture output value |
| *GCO* | grain crop output | output of grains |
| *OCO* | oil crop output | output of oil crops |
| *VO* | vegetable output | output of vegetables |
| *GFO* | garden fruit output | output of garden fruits |
| *PEO* | poultry egg output | output of poultry eggs |
| *TMO* | total meat output | output of total meat |
| *APO* | aquatic product output | output of aquatic products |

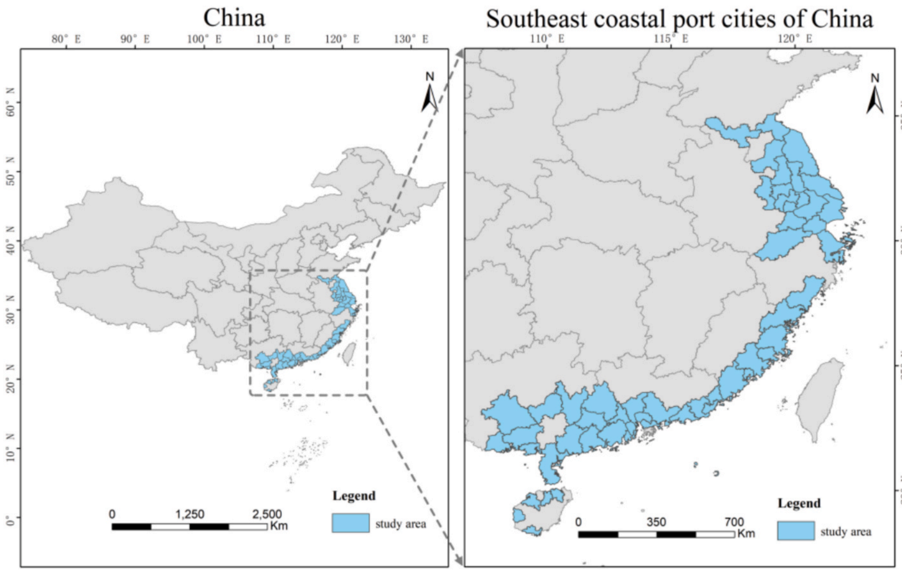


**Fig. 1.** Overview of the study area.

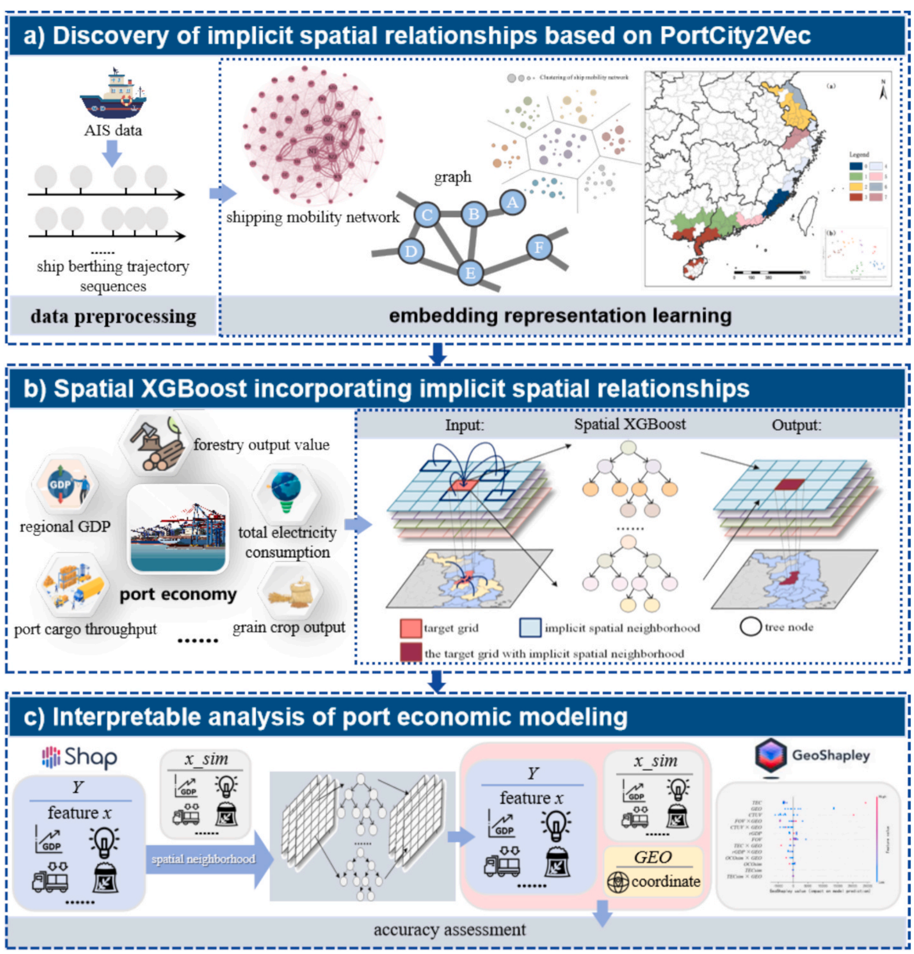


**Fig. 2.** Technical workflow.

continuous semantic vectors based on co-occurrence probabilities (Ascari et al., 2024)—we hypothesize that the co-occurrence of ships among port cities reflects deeper economic linkages. By embedding port cities with high co-occurrence probabilities into similar semantic vectors, we can reveal their implicit spatial relationships. Based on this, we propose the PortCity2Vec framework, which integrates AIS data with graph embedding techniques to identify and quantify implicit spatial relationships among port cities.

We construct implicit spatial relationships based on intercity interactions reflected in AIS data. These relationships do not rely on geographic proximity or administrative boundaries but are instead derived from network connections between port cities in terms of maritime accessibility and logistics flows, revealing potential patterns of economic interaction and spatial spillovers. Compared with explicit spatial structures, analyzing implicit spatial relationships offers several distinct advantages. While physical adjacency and administrative boundaries may indicate potential interaction, they often fail to reflect the reality of resource flows, information exchange, and economic activity. In contrast, implicit spatial relationships derived from AIS data capture how port cities are functionally connected through actual maritime behavior, rather than through mere physical closeness. This functional perspective reveals hidden spillovers, central-peripheral patterns, and cross-regional links often overlooked by conventional spatial frameworks. As a result, it provides a more accurate and insightful understanding of port city networks and the processes of regional integration.

First, we model the shipping mobility network as a directed weighted graph $G = (V, E, W)$, where $V$ denotes the set of port cities, $E$ represents directed edges indicating vessel movements, and $W$ indicates the edge weights, defined by the frequency of vessel mobility among port cities. Based on this network, we apply a random walk to generate sequences of port cities, aiming to uncover the implicit spatial relationships among them. The traditional DeepWalk method generates training sequences through uniform random walks (Xiong et al., 2022). While effective in general networks, it struggles to capture implicit spatial relationships in shipping mobility networks, which involve complex structures and heterogeneous nodes.

To address this issue, we introduce the biased random walk strategy to accurately capture the complex relationships among port cities. This strategy incorporates two adjustable parameters, $p$ and $q$, which control the direction and range of the walk, enabling a flexible balance between local and global structural information. This process is repeated to generate sequences of port cities for training samples. Specifically, when $p < 1$, the probability of returning to the previous node increases, which facilitates the exploration of physically adjacent port cities. Conversely, when $q < 1$, the walker is more likely to explore distant nodes, promoting cross-regional transitions and enhancing the model's ability to capture implicit spatial patterns. By applying biased random walks on the shipping mobility network, we generate port city sequences whose co-occurrence statistics are subsequently modeled by the Skip-gram framework. The exact sampling probabilities are given by Equation (1).

$$\pi_{i\to j} = \frac{\alpha_{pq}(t,j)W_{i\to j}}{\sum_k \alpha_{pq}(t,k)W_{i\to k}}\alpha_{pq}(t,j) = \begin{cases} \dfrac{1}{p}, d_{tj} = 0 \\ 1, d_{tj} = 1 \\ \dfrac{1}{q}, d_{tj} = 2 \end{cases} \tag{1}$$

Here: $\pi_{i\to j}$ is the transition probability from port city $i$ to port city $j$; $W_{i\to j}$ represents the shipping mobility frequency from port city $i$ to $j$; $t$ is the previous port city in the random walk; $d_{tj}$ denotes the shortest-path distance between $t$ and $j$; $\alpha_{pq}(t,j)$ is the bias factor in the biased random walk; $p$ and $q$ control the return and exploration tendencies of the walk, respectively.

As illustrated in Fig. 3a, the biased random walk proceeds from the previous node $t$ to the current node $i$. Fig. 3b shows the shipping mobility network among port cities, where the thickness of the arrows reflects the frequency of ship flows and node names are abbreviated. Fig. 3c illustrates the port city sequences generated through biased random walks on this network. For example, with a context window of 2 and a sequence length of 8, each node in the sequence is labeled with the abbreviation of its corresponding port city. Fig. 3d presents the resulting sample sequences. Based on these sequences, port cities are treated as "words" and sequences as "sentences." The Skip-gram model (Aka Uymaz and Kumova Metin, 2022; Zhang et al., 2024a) is then applied to learn embeddings that capture the implicit spatial relationships among

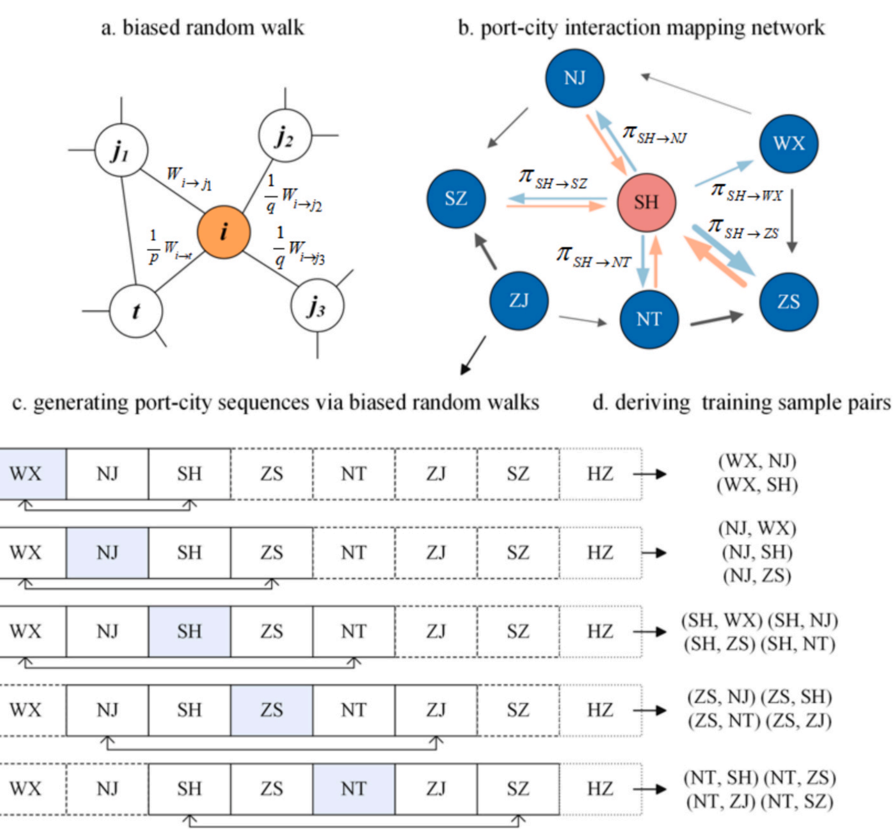


**Fig. 3.** Training sample deriving process (sampling window = 2, sequence length = 8, node name abbreviation of port city name).

port cities.

Fig. 4 shows the structure of the Skip-gram model. First, each port city is represented as a high-dimensional sparse vector through one-hot encoding and fed into the Input layer. Then, the model maps the high-dimensional one-hot encoding to a low-dimensional dense vector in the Projection layer by retrieving the corresponding row from the weight matrix, thus obtaining the embeddings of port cities. In this study, the embedding dimension was set to 128, based on iterative experiments. Experiments with dimensions of 64, 128, and 256 showed that 128-dimensional embeddings offered the most stable and discriminative representations for downstream tasks, including clustering and predictive modeling. This choice ensures that the learned embeddings effectively capture implicit spatial relationships among port cities while maintaining computational efficiency. In the Output layer, the model uses the embeddings to predict the surrounding port cities within the context window. The co-occurrence probabilities between the central port city and its context cities are dynamically computed through the softmax function. The model parameters are optimized by maximizing the prediction accuracy, thus generating more accurate port city embeddings. Equations (2)–(4) detail the core objective function and the embedding process.

$$v_w = x_w \bullet W_{pro} \tag{2}$$

$$p(w_o|w_c) = \frac{exp(v_{w_o} \bullet v_{w_c})}{\sum_{w=1}^{|V|} exp(v_w \bullet v_{w_c})} \tag{3}$$

$$\mathscr{L} = -\sum_{t=1}^{T} \sum_{-c \leq j \leq c, j \neq 0} logP(w_{t+j}|w_t) \tag{4}$$

Here, $x_w$ is the one-hot encoding of port city $w$, $W_{pro}$ is the projection weight matrix, and $v_w$ represents the embedding of port city $w$. $w_c$ and $w_o$ refer to the central and neighboring port cities, respectively, and $v_{w_c}$ and $v_{w_o}$ are their embeddings. |V| denotes the total number of port cities. $\mathscr{L}$ is the negative log-likelihood loss of the Skip-gram model. $T$ denotes the number of sampled sequences, and $c$ is the context window size. $w_t$ and $w_{t+j}$ represent the central port city and its neighboring port city within the context window, respectively. $P(w_{t+j}|w_t)$ denotes the predicted probability of observing $w_{t+j}$ in the context of $w_t$.

### 3.2. Spatial XGBoost model incorporating implicit spatial relationships

Traditional machine learning methods typically assume that data are independently and identically distributed (Wang et al., 2025), making it difficult to capture potential spatial dependencies, especially the implicit spatial relationships among port cities. In actual shipping mobility networks, connections among port cities are often not explicitly observable, but are embedded in complex interaction mechanisms and underlying economic relationships (Guo and Qin, 2022). As a result, conventional models struggle to make these relationships explicit, hindering a deeper understanding of port city interactions (Zhu et al., 2025). In view of this, we develop a spatial XGBoost model that incorporates port city embeddings, to account for such potential spatial dependencies and uncover hidden intercity connections.

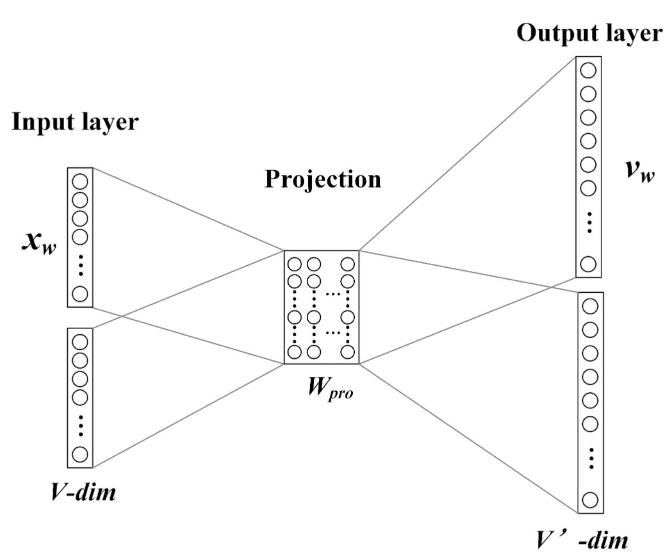


**Fig. 4.** Skip-gram model.

As an ensemble learning method based on gradient boosting, XGBoost offers high accuracy, strong generalization ability, and efficient handling of nonlinear features. It has been widely applied in socioeconomic analysis and spatial modeling (Zhang et al., 2024d). Unlike traditional models that only consider the socioeconomic indicators of individual port cities, our spatial XGBoost framework also integrates information from neighboring cities in the implicit spatial structure. Specifically, we incorporate socioeconomic indicators of both the target port city and its neighboring cities into the model as independent variables. This allows the model to capture not only city-specific characteristics but also the spatial context formed by their latent connections. As shown in Fig. 5, the spatial XGBoost model extends the traditional framework by incorporating features from both the target city and its implicit neighbors, enhancing its ability to model complex economic interactions.

Port cargo throughput, a key indicator of port economic activity, reflects not only regional economic development, but also logistics efficiency and industrial agglomeration. We use port cargo throughput as the dependent variable and integrate the socioeconomic indicators of both the target port city and its neighbors in the implicit spatial structure as independent variables to capture complex economic interactions more effectively. Table 3 lists the specific indicators used in the modeling, covering four dimensions: economic foundation, logistics capacity, industrial characteristics, and spatial association. Specifically, these include regional GDP, total social electricity consumption, cargo transportation volume, cargo turnover, grain crop output, aquatic product output, etc. Regional GDP and electricity consumption indicate the economic activity in port cities, serving as key measures of the economic foundation. Cargo transportation volume and turnover reflect port logistics efficiency, while grain crop and aquatic product outputs highlight the industrial structure and key import-export characteristics. In terms of spatial association, we build an implicit spatial structure based on port city embeddings, and incorporate neighboring cities' socioeconomic indicators, enabling the model to capture spatial dependencies and better represent implicit spatial relationships.

### 3.3. Interpretability analysis of port economic modeling

The traditional SHAP methods mainly focus on the marginal contributions of individual features to the model, which makes it difficult to capture complex spatial relationships (Hu et al., 2024). GeoShapley incorporates geographic features into the interpretability framework of machine learning models, revealing complex interaction mechanisms in regional economies (Liu, 2024). It also provides a theoretical foundation for addressing the limitations of the independence assumption in traditional SHAP methods when applied to spatial data. Accordingly, we apply the GeoShapley method to a machine learning model that accounts for implicit spatial relationships, aiming to better understand the driving mechanisms of the port economy.

The GeoShapley method extends the classic SHAP framework by integrating spatial weight matrices, geographic locations, or spatial adjacency information to uncover how geographic features influence model predictions (Li, 2024). Specifically, GeoShapley extends the classical SHAP formula by incorporating a spatial weight matrix $W$ and modifies the prediction function to reflect spatial influence. The core

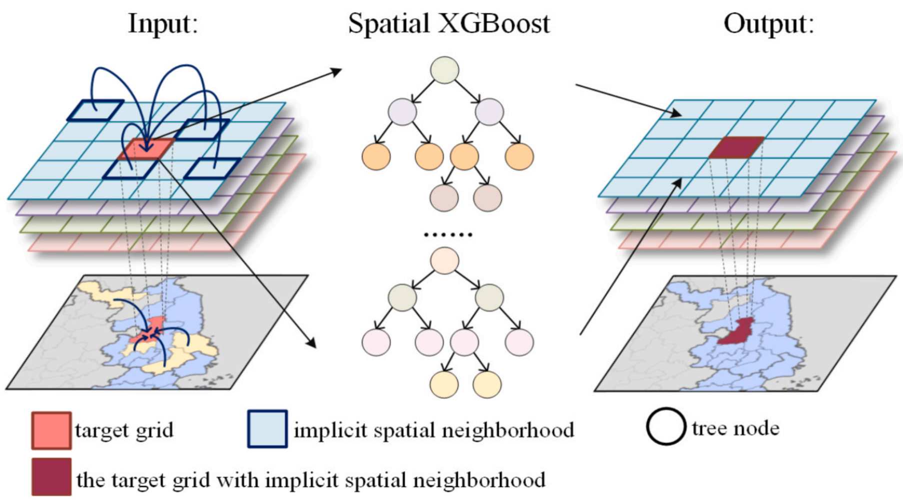


**Fig. 5.** Schematic diagram of spatial XGBoost.

**Table 3**
Each dimension and specific indicators.

| Dimension | Indicator |
|---|---|
| economic foundation | regional GDP; total output value of agriculture, forestry, animal husbandry, and fishery; agricultural output value; forestry output value; animal husbandry output value; fishery output value; total electricity consumption; industrial electricity consumption |
| logistics capacity | cargo transportation volume; cargo turnover volume |
| industrial characteristic | grain crop output; oil crop output; vegetable output; garden fruit output; poultry egg output; total meat output; aquatic product output |
| spatial association | the same socioeconomic indicators of neighboring port cities in the implicit spatial structure |

computation is shown in Equation (5).

$$\varphi_j^{geo} = \sum_{SF\setminus\{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!}[f(S\cup\{j\},W) - f(S,W)] \tag{5}$$

Here, $\varphi_j^{geo}$ denotes the Shapley contribution value of feature $j$ under the GeoShapley method. $S$ is a subset of the full feature set $F$. $W$ represents the spatial weight matrix. The term $f(S,W)$ refers to the model prediction rbased on the feature subset $S$ while incorporating the spatial weights $W$. The difference $f(S\cup\{j\},W) - f(S,W)$ captures the marginal contributions of feature $j$ to the model prediction, with spatial dependencies taken into account.

The introduction of the spatial weight matrix W means that the contribution of each feature depends not only on the feature values of the target port city itself, but also on those of neighboring port cities. This allows for a more accurate reflection of cross-regional economic relationships and spatial dependencies.

## 4. Experimental analysis

### 4.1. Implicit spatial relationship among port cities

This study focuses on port cities along the southeastern coast of China to explore the potential economic connections among port cities through embedding representation learning. Therefore, we construct the shipping mobility network based on AIS data. As shown in Fig. 6, the network visually illustrates the interaction patterns among port cities.

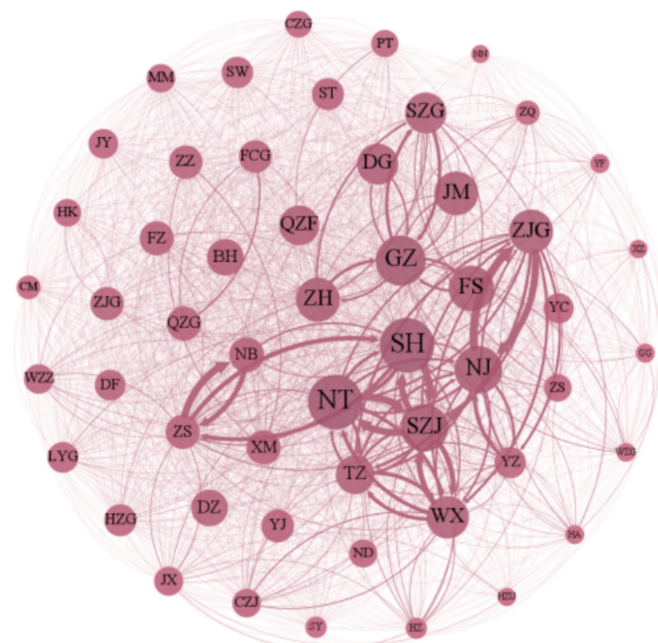


**Fig. 6.** Shipping ability network.

To uncover the latent connections among port cities in a high-dimensional space, PortCity2Vec is applied to generate embeddings, and cosine similarity is used to measure intercity similarity. Fig. 7 presents a heatmap of similarity scores calculated based on the embeddings, revealing clear clustering patterns among the port cities. Port cities in the Yangtze River Delta, the Pearl River Delta and its southwestern inland area, as well as the Beibu Gulf–Hainan region, form distinct spatial clusters. In addition, port cities in the Minnan region and eastern Guangdong also exhibit agglomeration, highlighting geographic continuity and regional integration. Meanwhile, port cities such as Chaozhou, Lianyungang, and Yancheng display clustering patterns that differ from the major regional groups in the similarity heatmap. This phenomenon of non-physical adjacency clustering indicates that the potential connections among port cities are not entirely constrained by physical distance. Such non-physical adjacency clustering patterns reflect the advantage of PortCity2Vec in revealing potential connections among port cities. Compared to conventional methods that rely on physical adjacency or administrative boundaries, PortCity2Vec enables the discovery of cross-regional relationships beyond these physical constraints.

Guided by the biased random walk, PortCity2Vec captures the

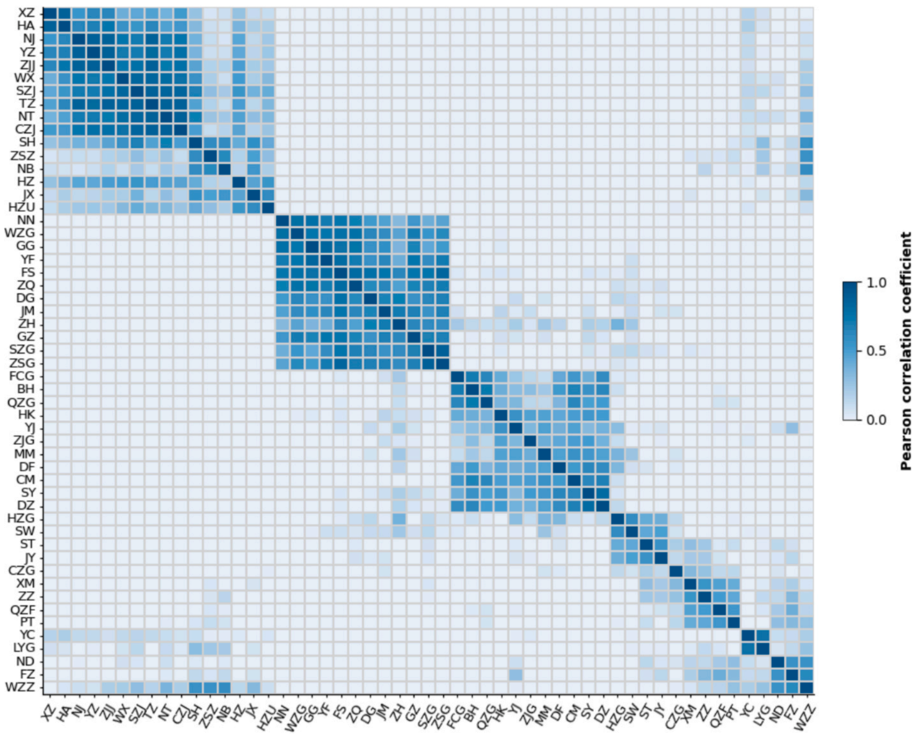


**Fig. 7.** Similarity heatmap of port city embeddings.

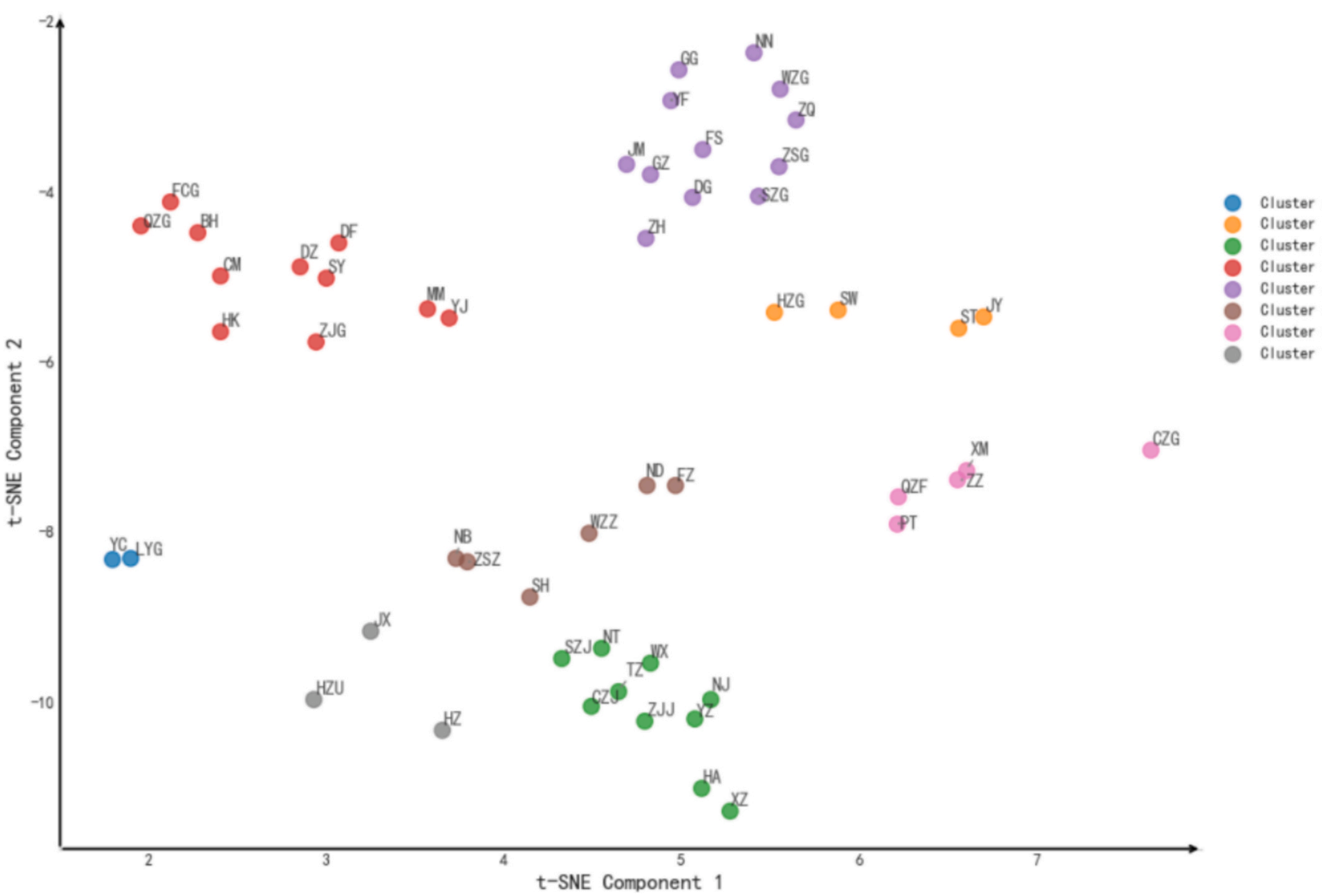


**Fig. 8.** Port-city embeddings 2D clustering map.

structural and neighborhood features of nodes in the network, generating a 128-dimensional embedding for each port city as its representation in the shipping mobility network. To extract embeddings more effectively, the dimensionality was first reduced from 128 to 50 using PCA (Principal Component Analysis) to preserve the global structure. Then, t-SNE (t-distributed Stochastic Neighbor Embedding) was applied to further reduce the dimensionality to two, emphasizing local proximity relationships. As shown in Fig. 8, the clustering results reveal potential connections among port cities in the shipping mobility network.

Fig. 9 shows the spatial distribution of the clustering results based on port city embeddings. Port cities such as Xiamen, Quanzhou, Putian, Zhangzhou, and Chaozhou are grouped into the same cluster, reflecting geographic proximity and regional integration. Port cities along the Yangtze River—such as Nanjing, Suzhou, Zhenjiang, and Nantong—form a cluster despite belonging to different administrative regions, driven by frequent vessel movement along the main river channel and their shared role as key inland shipping hubs. A clear pattern of non-physical adjacency clustering is also evident in the embedding representation. For example, Shanghai, Zhoushan, Ningbo, Wenzhou, Fuzhou, and Ningde, despite spanning multiple provinces, share highly similar functional connections within the network. Similarly, Zhuhai, Shenzhen, Guangzhou, Nanning, and Guigang—though geographically distant across Guangdong and Guangxi—share notable similarities due to their strategic locations at key shipping intersections. The Beibu Gulf–Hainan region also exhibits representative clustering characteristics. Despite being geographically discontinuous and separated by sea, port cities such as Qinzhou, Beihai, Zhanjiang, Haikou, and Sanya exhibit strong functional coordination and maritime connectivity within the shipping mobility network. These patterns indicate significant non-physical adjacency clustering among port cities in the embedding representation. Such clustering goes beyond traditional notions of physical proximity and more accurately reflects the potential connections among port cities.

### 4.2. *Interpretable port economic model considering embedding representation*

To evaluate the impact of different feature combinations on the prediction of port cargo throughput, we designed three modeling schemes based on the XGBoost model for comparative analysis: (1) Baseline model: Uses only the socioeconomic indicators of the target port city. (2) Implicit spatial structure enhanced model: Incorporates the socioeconomic indicators of neighboring cities identified through the implicit spatial structure. (3) Spatial embedding enhancement model:

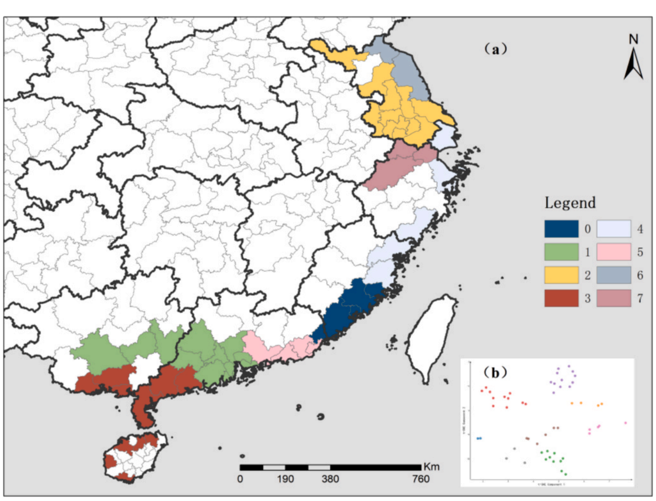


**Fig. 9.** Spatial distribution of port city clustering results.

Builds on the previous model (2) by further incorporating a geographic embedding feature (*GEO*) based on the latitude and longitude coordinates of port cities. The modeling results for each scheme are presented in Table 4.

The baseline model, using only the target city's socioeconomic indicators, achieves an $R^2$ of 0.6554, explaining about 66% of the variance but with limited predictive capability. The model incorporating the implicit spatial structure demonstrates a notable improvement, with the $R^2$ rising to 0.8309, suggesting that features from similar port cities help capture potential economic connections. The spatial embedding enhancement model, which incorporates *GEO*, further improves performance, achieving an $R^2$ of 0.8541 and maintaining a high level of predictive accuracy.

Fig. 10 presents the SHAP value plot of the baseline model, illustrating both the strength and direction of each feature's impact on the prediction outcomes. In the plot, color indicates the magnitude of the feature values (red for higher and blue for lower), while the sign of the SHAP values reveals whether the feature contributes positively or negatively to the model's dependent variable. As shown in the figure, *CTUV* and *TEC* exhibit the widest range of SHAP value contributions, indicating that they are the most influential factors in predicting port cargo throughput. Notably, *CTUV* shows a strong positive correlation with *Y*, suggesting that higher transport demand is associated with stronger port handling capacity. In addition, *GCO*, *FIOV*, and *APO* positively impact *Y*, while *FOV*, *AOV*, and *TMO* show negative correlations.

In summary, port cargo throughput is shaped by multiple socioeconomic factors, and each factor shows a distinct pattern of influence.

The implicit spatial structure enhanced model shows significantly improved explanatory power, revealing regional connections and spatial spillover effects in port transportation. Fig. 11 shows that *CTUV* and *TEC* still rank among the top in terms of SHAP values, remaining the core drivers of port cargo throughput. Notably, the importance of *FOV* has significantly increased, now ranking third in SHAP values and emerging as a key variable. *CTUV*, *GCO*, and *FIOV* still exhibit stable and significant positive contributions. The *TEC* of similar cities (*TECsim*) has also become a key variable, showing a positive correlation with *Y* and highlighting the interconnectedness of cross-regional economic activities. The industrial activities of nearby cities or those within the implicit spatial structure have spillover effects on local port transportation activities. Compared to other factors, *FOV* and *OCO* show a negative correlation with *Y*, indicating a potential mismatch between these industries and port transportation. Additionally, the importance of *rGDP* in the model is relatively low, indicating its limited explanatory power for port transportation, which deviates from traditional perceptions. While *rGDP* is often regarded as a key driver of port transportation, the empirical results suggest its direct impact is limited.

In summary, incorporating the socioeconomic indicators of the implicit spatial neighborhood significantly reveals spatial spillover effects, underscoring the importance of cross-regional economic cooperation in port transportation.

Building upon the implicit spatial structure enhanced model, we further incorporated *GEO*, which improved the model's capacity to recognize the spatial patterns of port cargo throughput. As shown in Fig. 12, the results indicate that *TEC* is the most influential factor among

**Table 4**
Results of model schemes.

| Model scheme | Original feature number | Feature number after screening | $R^2$ |
|---|---|---|---|
| Baseline model | 17 | 13 | 0.6554 |
| Implicit spatial structure enhanced model | 34 | 12 | 0.8309 |
| Spatial embedding enhancement model | 35 | 7 | 0.8541 |

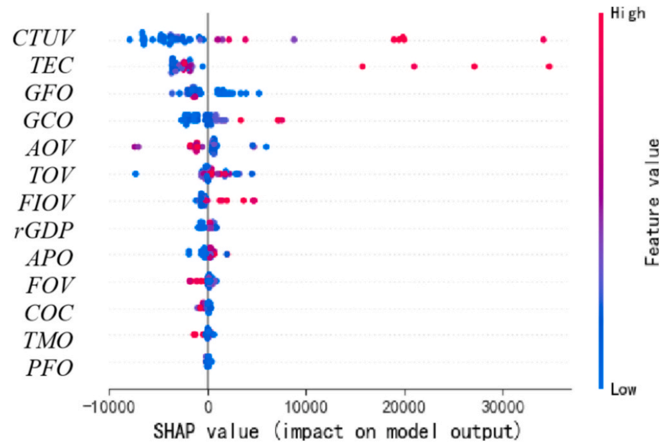


**Fig. 10.** SHAP value plot using only socioeconomic indicators of target city.

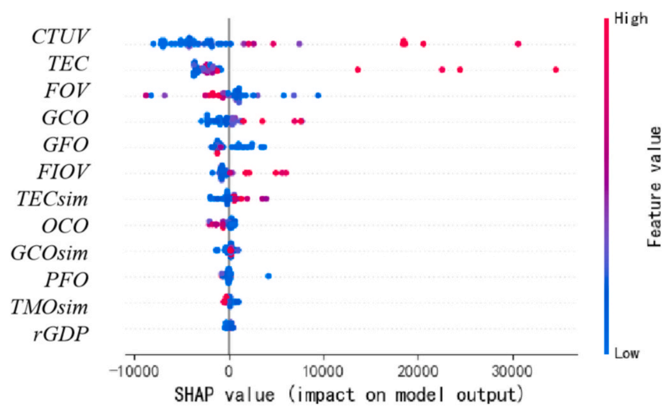


**Fig. 11.** SHAP value plot incorporating socioeconomic indicators of neighboring city.

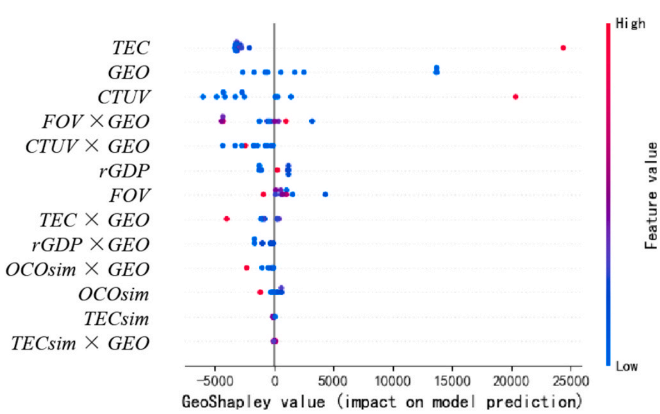


**Fig. 12.** GeoShapley value plot incorporating GEO and neighboring city.

all variables in the spatial embedding enhancement model. It also exhibits a significant positive nonlinear impact on *Y*. *GEO* ranks just behind *TEC*, highlighting its crucial role in the model. *CTUV* remains positively correlated with *Y*, while *FOV* continues to show a negative correlation. As a proxy of macroeconomic scale, *rGDP* still demonstrates relatively limited explanatory power, reinforcing that port logistics growth is not only driven by *rGDP*. In addition, differences in economic structure among port cities also play a moderating role. Moreover, analyzing the interaction between *GEO* and key variables reveals the mechanisms behind regional spatial heterogeneity in port cargo throughput, further highlighting the critical role of spatial relationships in port transportation. Specifically, the interaction between *FOC* and *GEO* is nonlinear, indicating that its impact on port transportation is influenced not only by the scale of output but also by *GEO*. The interaction between *CTUV* and *GEO* shows a generally positive correlation, suggesting that more active cargo flow provides stronger support for port operations. However, geographic disadvantages may constrain its full potential. In contrast, the interaction between *TEC* and *GEO* is negatively correlated, implying that in geographically disadvantaged port cities, frequent economic activity has a diminished impact on port transportation.

In summary, *GEO* significantly improves the model's ability to identify the geographical advantages of port cities. As key representations of implicit spatial constraints, geographic features help compensate for the limitations of socioeconomic indicators in capturing spatial dependencies, highlighting the importance of incorporating implicit spatial relationships into the model.

## 5. Discussion

Although existing studies have made progress in exploring the economic connections of port cities, most of them rely on static statistical indicators or physical proximity. As a result, existing methods often overlook connections between port cities that are not physically adjacent but exhibit functional similarity in the implicit spatial structure. Moreover, traditional spatial modeling approaches struggle to capture high-dimensional spatial relationships and lack interpretability mechanisms, limiting deeper insights into port city economic connections. To address these issues, we propose the PortCity2Vec framework, which identifies port cities that are not physically adjacent but are close in the implicit spatial structure, based on co-occurrence probabilities in the shipping mobility network. This framework generates high-dimensional embeddings capable of capturing potential economic connections. By integrating the socioeconomic indicators of similar port cities with those of the target port city, we construct a spatial XGBoost model that accounts for implicit spatial dependencies. We quantify the marginal contributions of each variable and geographic features using the GeoShapley method, enhancing model interpretability. This study overcomes the limitations of traditional analysis based on physical adjacency, reveals latent economic connections, and highlights the key role of spatial embedding techniques in identifying potential connections among port cities.

Before discussing the empirical implications, it should be emphasized that the implicit spatial relationships identified in this study are derived specifically from shipping interactions captured by AIS data. Therefore, the findings primarily reflect shipping-oriented connectivity and economic connections among port cities. More importantly, the concept of implicit spatial relationships proposed in this study is not confined to shipping networks. It reflects a broader spatial interaction paradigm in which functional proximity is inferred from mobility co-occurrence patterns rather than physical adjacency or administrative contiguity. From a methodological perspective, the PortCity2Vec framework provides a transferable approach for identifying latent spatial structures in other transportation and mobility systems. For instance, in aviation networks, flight trajectory data or passenger flow records could be used to uncover implicit spatial relationships among airports that are functionally connected but physically distant. Similarly, in road freight systems, truck trajectory data or logistics flow information may reveal potential economic connections between cities that are not captured by traditional distance-based or adjacency-based spatial models. Consequently, while this study focuses on shipping-oriented port city interactions, the proposed framework offers a generalizable analytical paradigm for modeling implicit spatial dependencies in a wide range of mobility-driven economic systems.

This study complements and reinforces existing literature both theoretically and empirically by uncovering the complex economic interactions among port cities. The findings show that port cargo throughput is significantly correlated with variables such as cargo turnover volume and grain crop output. This aligns with the study by Li et al. (2019), who used a system dynamics model to examine green coordinated development at Shanghai Port and emphasized the central role of cargo flow in port economic activity. These results suggest that, whether using machine learning models or system dynamics approaches, cargo flow intensity remains a robust and reliable indicator for characterizing port economic growth. Furthermore, the negative correlation between forestry output and port cargo throughput suggests a structural mismatch between regional industrial profiles and port functions. In regions dominated by forestry, high production levels do not necessarily translate into high port throughput, indicating that port infrastructure and supporting systems may not be well aligned with local economic activities. This interpretation is consistent with recent evidence showing that port performance depends more on the coherence of the broader port-industry system than on industrial scale alone, and that system-wide economic dynamics and hinterland restructuring can lead to uneven port outcomes even in economically active regions, such as those observed in the Pearl River Delta (Zhang et al., 2025; Li et al., 2022). In terms of spatial mechanism analysis, this study identifies non-physically adjacent relationships among port cities through the PortCity2Vec framework and the GeoShapley method, moving beyond the traditional focus on physical adjacency. This finding echoes Gaidelys and Benetyte (2021), who emphasized that interactions among port cities do not solely rely on physical adjacency. It is worth noting that although GDP has long been a central focus in studies examining the relationship between economic growth and transportation in port cities. Li et al. (2019) and Liu et al. (2022), for instance, regarded GDP as a key indicator of port city development. However, our empirical findings suggest that GDP contributes little to explaining variations in port cargo throughput. This contrasts with Li et al. (2019), who identified GDP as the primary driver of port city growth. It also aligns with Nogué-Algueró (2020), who criticized the GDP-centric mindset for overlooking issues such as uneven resource allocation, industrial imbalance, and functional differentiation.

Although this study contributes to methodological innovation and empirical analysis, several limitations remain. First, while the research focuses on modeling complex spatial interactions among port cities, the construction of feature variables remains limited. Future work could incorporate multi-source heterogeneous data, such as enterprise registration records, mobile phone signaling data, and remote sensing observations. This would allow a more comprehensive capture of the functional characteristics and economic interactions of port cities. This would help enhance the model's explanatory power. Second, the relatively short temporal scale adopted in this study constrains the ability to capture the evolving nature of economic relationships among port cities. Future research could leverage longer time-series data and apply spatiotemporal analytical methods to explore these dynamic patterns. Third, integrating multi-modal transportation data—including road, rail, and air networks alongside maritime shipping—would provide a more holistic understanding of port city economic systems. While shipping remains the primary mode for port cities, other forms of transportation also play a significant role in economic and spatial integration. Therefore, relying solely on shipping networks may fail to fully capture the comprehensive connectivity and interactions among port cities.

## 6. Conclusion

As ports play an increasingly important role in regional economic integration, the port economy has become a cornerstone of regional coordinated development. Accordingly, exploring the driving mechanisms behind this process is of particular importance. In the era of big data, the traditional analytical paradigms based on statistical econometrics struggle to effectively capture the complex spatial interactions underlying port economies.

In response to the dual demands in current research for explicit spatial modeling and interpretability, we propose an analytical framework that integrates embedding representation learning with interpretable machine learning to capture implicit spatial structures. We systematically explore the economic connections and implicit spatial structure of port cities along the southeastern coast of China. The study draws the following conclusions: (1) There are implicit spatial relationships among port cities that go beyond physical adjacency. Certain port cities, including Chaozhou, Lianyungang, and Yancheng, form clusters that are not based on physical adjacency, highlighting the role of implicit spatial connections in port economic interactions. (2) Introducing socioeconomic indicators and geographic features of neighboring cities within the implicit spatial structure effectively improves the accuracy of the port economic model. Compared to using only local socioeconomic indicators, the $R^2$ increased from 0.6554 to 0.8541. (3) The development of the port economy is driven by multiple factors. A strong positive correlation exists between port cargo throughput and variables such as cargo turnover volume and grain crop output, which reflects the importance of economic activity and transportation intensity. Conversely, negative correlations with agricultural and forestry output suggest possible mismatches between these resource sectors and port functions. Regional GDP shows limited contribution, implying it is not a decisive factor in port economic development. (4) By integrating implicit spatial relationships identified through PortCity2Vec with geographic embedding features, this approach effectively overcomes limitations of traditional models. It significantly enhances both model performance and interpretability, allowing for more precise identification of potential spatial interactions among port cities.

In summary, this study captures implicit spatial relationships among port cities by integrating embedding-based representation learning with an interpretable machine learning framework. By moving beyond physical adjacency and explicitly observed connections, the proposed approach reveals implicit connections embedded in shipping mobility networks, thereby overcoming key limitations of traditional spatial econometric analyses. The findings enrich the theoretical understanding of port economic interaction mechanisms and provide quantitative support for regional coordinated development and port planning strategies. It should be emphasized that the empirical analysis is grounded in AIS data and thus reflects shipping-centered mechanisms of port economic interactions. However, the analytical logic of modeling implicit spatial relationships is not confined to maritime contexts. With appropriate mobility or trajectory data, the framework can be extended to aviation networks to uncover latent relationships among airports, or to road-based freight systems to reveal implicit intercity logistics interactions. Therefore, the proposed framework provides a flexible and transferable analytical tool for studying spatial interactions across diverse transportation modes and economic systems.

**Funding**.

This project was supported by National Key Research and Development Program of China under grant 2023YFB3906804.

## CRediT authorship contribution statement

**Xiaoli Cai:** Writing – original draft, Methodology, Formal analysis. **Sheng Wu:** Writing – original draft, Supervision, Resources, Project administration, Conceptualization. **Peixiao Wang:** Writing – original draft, Supervision, Software, Methodology, Conceptualization. **Hengcai Zhang:** Writing – review & editing, Writing – original draft, Software. **Shifen Cheng:** Writing – review & editing, Writing – original draft, Methodology. **Feng Lu:** Writing – review & editing, Writing – original draft, Methodology.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data supporting the main findings of this study are available at https://doi.org/10.6084/m9.figshare.29222318.

## References


Aka Uymaz, H., Kumova Metin, S., 2022. Vector based sentiment and emotion analysis from text: a survey. Eng. Appl. Artif. Intel. 113, 104922. https://doi.org/10.1016/j.engappai.2022.104922.

Ascari, R., Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., 2024. A Fistful of Vectors: a Tool for Intrinsic Evaluation of Word Embeddings. Cogn. Comput. 16 (3), 949–963. https://doi.org/10.1007/s12559-023-10235-3.

Bhattarai, P., Thakuri, D.S., Nie, Y., Chand, G.B., 2024. Explainable AI-based Deep-SHAP for mapping the multivariate relationships between regional neuroimaging biomarkers and cognition. Eur. J. Radiol. 174, 111403. https://doi.org/10.1016/j.ejrad.2024.111403.

Černevičienė, J., Kabašinskas, A., 2024. Explainable artificial intelligence (XAI) in finance: a systematic literature review. Artif. Intell. Rev. 57 (8), 216. https://doi.org/10.1007/s10462-024-10854-8.

Chen, X., Qu, R., Dong, J., Dong, H., Bai, R., 2024. Advancing container port traffic simulation: a data-driven machine learning approach in sparse data environments. Appl. Soft Comput. 166, 112190. https://doi.org/10.1016/j.asoc.2024.112190.

Cong, L., Zhang, D., Wang, M., Xu, H., Li, L., 2020. The role of ports in the economic development of port cities: Panel evidence from China. Transp. Policy 90, 13–21. https://doi.org/10.1016/j.tranpol.2020.02.003.

Durán, C., Fernández-Campusano, C., Carrasco, R., Carrillo, E., 2024. DMLBC: Dependable machine learning for seaports using blockchain technology. J. King Saud Univ.-Comput. Inf. Sci. 36 (1), 101918. https://doi.org/10.1016/j.jksuci.2024.101918.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., Ranjan, R., 2023. Explainable AI (XAI): Core ideas, Techniques, and Solutions. ACM Comput. Surv. 55 (9), 1–33. https://doi.org/10.1145/3561048.

Feng, H., Lin, Q., Zhang, X., Lam, J.S.L., Yap, W.Y., 2024. Port selection by container ships: a big AIS data analytics approach. Res. Transp. Bus. Manag. 52, 101066. https://doi.org/10.1016/j.rtbm.2023.101066.

Filom, S., Amiri, A.M., Razavi, S., 2022. Applications of machine learning methods in port operations – a systematic literature review. Transp. Res. Pt. e-Logist. Transp. Rev. 161, 102722. https://doi.org/10.1016/j.tre.2022.102722.

Gaidelys, V., Benetyte, R., 2021. Analysis of the Competitiveness of the Performance of Baltic ports in the Context of Economic Sustainability. Sustainability 13 (6), 3267. https://doi.org/10.3390/su13063267.

Ghebrehiwet, I., Zaki, N., Damseh, R., Mohamad, M.S., 2024. Revolutionizing personalized medicine with generative AI: a systematic review. Artif. Intell. Rev. 57 (5), 128. https://doi.org/10.1007/s10462-024-10768-5.

Guo, J., Qin, Y., 2022. Coupling characteristics of coastal ports and urban network systems based on flow space theory: Empirical evidence from China. Habitat Int. 126, 102624. https://doi.org/10.1016/j.habitatint.2022.102624.

Henríquez, R., Martínez De Osés, F.X., Martínez Marín, J.E., 2022. Technological drivers of seaports' business model innovation: an exploratory case study on the port of Barcelona. Res. Transp. Bus. Manag. 43, 100803. https://doi.org/10.1016/j.rtbm.2022.100803.

Hu, X., Zhu, M., Feng, Z., Stanković, L., 2024. Manifold-based Shapley explanations for high dimensional correlated features. Neural Netw. 180, 106634. https://doi.org/10.1016/j.neunet.2024.106634.

Ishikawa, S., Todo, M., Taki, M., Uchiyama, Y., Matsunaga, K., Lin, P., Ogihara, T., Yasui, M., 2023. Example-based explainable AI and its application for remote sensing image classification. Int. J. Appl. Earth Obs. Geoinf. 118, 103215. https://doi.org/10.1016/j.jag.2023.103215.

Li, K.X., Li, M., Zhu, Y., Yuen, K.F., Tong, H., Zhou, H., 2023. Smart port: a bibliometric review and future research directions. Transp. Res. Pt. e-Logist. Transp. Rev. 174, 103098. https://doi.org/10.1016/j.tre.2023.103098.

Li, S., Haralambides, H., Zeng, Q., 2022. Economic forces shaping the evolution of integrated port systems - the case of the container port system of China's Pearl River Delta. Res. Transp. Econ. 94, 101183. https://doi.org/10.1016/j.retrec.2022.101183.

Li, Y., Zhang, X., Lin, K., Huang, Q., 2019. The Analysis of a simulation of a Port–City Green Cooperative Development, based on System Dynamics: a Case Study of Shanghai Port. China. Sustainability 11 (21), 5948. https://doi.org/10.3390/su11215948.

Li, Z., 2024. GeoShapley: a Game Theory Approach to measuring Spatial Effects in Machine Learning Models. Ann. Am. Assoc. Geogr. 114 (7), 1365–1385. https://doi.org/10.1080/24694452.2024.2350982.

Lim, S., Kim, S.J., Park, Y., Kwon, N., 2021. A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic. Expert Syst. Appl. 184, 115532. https://doi.org/10.1016/j.eswa.2021.115532.

Lin, B., Tang, J., Dai, C., Lu, B., 2024. Spatial effect of digital economy on the coordinated development of Port cities. Res. Transp. Econ. 103, 101385. https://doi.org/10.1016/j.retrec.2023.101385.

Liu, C., Liu, J., Zhou, X., Zhao, Z., Wan, C., Liu, Z., 2020. AIS data-driven approach to estimate navigable capacity of busy waterways focusing on ships entering and leaving port. Ocean Eng. 218, 108215. https://doi.org/10.1016/j.oceaneng.2020.108215.

Liu, L., 2024. An ensemble framework for explainable geospatial machine learning models. Int. J. Appl. Earth Obs. Geoinf. 132, 104036. https://doi.org/10.1016/j.jag.2024.104036.

Liu, R.W., Zhou, S., Liang, M., Gao, R., Wang, H., 2024. From ports to routes: Extracting multi-scale shipping networks using massive AIS data. Ocean Eng. 311, 118969. https://doi.org/10.1016/j.oceaneng.2024.118969.

Liu, W., Yang, Y., Luo, Q., Zeng, X., Chen, C., Zhu, J., Lin, W., Chen, H., Huo, W., He, M., Jin, Y., Xu, J., 2022. Study on the Contribution of Seaport to Urban Economy: an Empirical and Quantitative Analysis of Xiamen Port. J. Mar. Sci. Eng. 10 (11), 1753. https://doi.org/10.3390/jmse10111753.

Nogué-Algueró, B., 2020. Growth in the docks: ports, metabolic flows and socio-environmental impacts. Sustain. Sci. 15 (1), 11–30. https://doi.org/10.1007/s11625-019-00764-y.

Nohara, Y., Matsumoto, K., Soejima, H., Nakashima, N., 2022. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput. Meth. Programs Biomed. 214, 106584. https://doi.org/10.1016/j.cmpb.2021.106584.

Park, S., Yang, J.-S., 2022. Interpretable deep learning LSTM model for intelligent economic decision-making. Knowledge-Based Syst. 248, 108907. https://doi.org/10.1016/j.knosys.2022.108907.

Peng, P., Cheng, S., Lu, F., 2020. Characterizing the global liquefied petroleum gas trading community using mass vessel trajectory data. J. Clean. Prod. 252, 119883. https://doi.org/10.1016/j.jclepro.2019.119883.

Peng, P., Claramunt, C., Cheng, S., Yang, Y., Lu, F., 2023. A multi-layer modelling approach for mining versatile ports of a global maritime transportation network. Int. J. Digit. Earth 16 (1), 2129–2151. https://doi.org/10.1080/17538947.2023.2220614.

Peng, P., Yang, Y., Lu, F., Cheng, S., Mou, N., Yang, R., 2018. Modelling the competitiveness of the ports along the Maritime Silk Road with big data. Transp. Res. Pt. A-Policy Pract. 118, 852–867. https://doi.org/10.1016/j.tra.2018.10.041.

Ueda, K., Suwa, H., Yamada, M., Ogawa, Y., Umehara, E., Yamashita, T., Tsubouchi, K., Yasumoto, K., 2024. SSCDV: Social media document embedding with sentiment and topics for financial market forecasting. Expert Syst. Appl. 245, 122988. https://doi.org/10.1016/j.eswa.2023.122988.

Wang, P., Zhang, H., Zhang, B., Cheng, S., Lu, F., Zhang, T., 2025. Spatiotemporal variations of private e-bike trips with explainable data-driven technologies. Cities 158, 105712. https://doi.org/10.1016/j.cities.2025.105712.

Wang, R., Tan, Q., 2019. Dynamic Model of Port Throughput's Influence on Regional Economy. J. Coast. Res. 93 (sp1), 811. https://doi.org/10.2112/SI93-114.1.

Xiong, H., Yan, J., Huang, Z., 2022. Learning Regularized Noise Contrastive Estimation for Robust Network Embedding. IEEE Trans. Knowl. Data Eng. 1–1. https://doi.org/10.1109/TKDE.2022.3148284.

Xu, Y., Peng, P., Lu, F., Claramunt, C., 2024. Uncovering the multiplex network of global container shipping: Insights from shipping companies. J. Transp. Geogr. 120, 103991. https://doi.org/10.1016/j.jtrangeo.2024.103991.

Yang, D., Wu, L., Wang, S., Jia, H., Li, K.X., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. Transp. Rev. 39 (6), 755–773. https://doi.org/10.1080/01441647.2019.1649315.

Yen, B.T.H., Huang, M.-J., Lai, H.-J., Cho, H.-H., Huang, Y.-L., 2023. How smart port design influences port efficiency – a DEA-Tobit approach. Res. Transp. Bus. Manag. 46, 100862. https://doi.org/10.1016/j.rtbm.2022.100862.

Zhang, H., Kou, G., Peng, Y., Zhang, B., 2024a. Role-aware random walk for network embedding. Inf. Sci. 652, 119765. https://doi.org/10.1016/j.ins.2023.119765.

Zhang, J., Yang, D., Luo, M., 2024b. Port efficiency types and perspectives: a literature review. Transp. Policy 156, 13–24. https://doi.org/10.1016/j.tranpol.2024.07.014.

Zhang, J.-Q., Woo, S.-H., Li, K.X., 2024c. Port-city synergism and regional development policy: evidence from the Yangtze River Region. Transp. Res. Pt. e-Logist. Transp. Rev. 192, 103817. https://doi.org/10.1016/j.tre.2024.103817.

Zhang, Q., Hu, X., Li, Z., Zhou, Y., Qi, X., 2025. Port's industry ecosystem construction: Empirical evidence from China. Transp. Policy 163, 42–60. https://doi.org/10.1016/j.tranpol.2025.01.005.

Zhang, T., Yin, J., Wang, X., Min, J., 2024d. Prediction of container port congestion status and its impact on ship's time in port based on AIS data. Marit. Policy Manag. 51 (5), 669–697. https://doi.org/10.1080/03088839.2023.2165185.

Zhao, D., Zhen-fu, L., Yu-tao, Z., Xiao, C., Shan-shan, L., 2020. Measurement and spatial spillover effects of port comprehensive strength: Empirical evidence from China. Transp. Policy 99, 288–298. https://doi.org/10.1016/j.tranpol.2020.09.006.

Zhou, X., Huang, Z., Xia, T., Zhang, X., Duan, Z., Wu, J., Zhou, G., 2025. The integrated application of big data and geospatial analysis in maritime transportation safety management: a comprehensive review. Int. J. Appl. Earth Obs. Geoinf. 138, 104444. https://doi.org/10.1016/j.jag.2025.104444.

Zhou, Y., Yuen, K.F., 2024. Prepare for the sustainability era: a quantitative risk analysis model for container shipping sustainability-related risks. J. Clean. Prod. 475, 143661. https://doi.org/10.1016/j.jclepro.2024.143661.

Zhu, M., Cao, J., Zheng, X., Huang, Z., Chen, S., Wen, Y., 2025. Predicting multi-port vessel traffic flow: an improved spatial-temporal graph neural network with uncertainty quantification. Ocean Eng. 329, 121135. https://doi.org/10.1016/j.oceaneng.2025.121135.